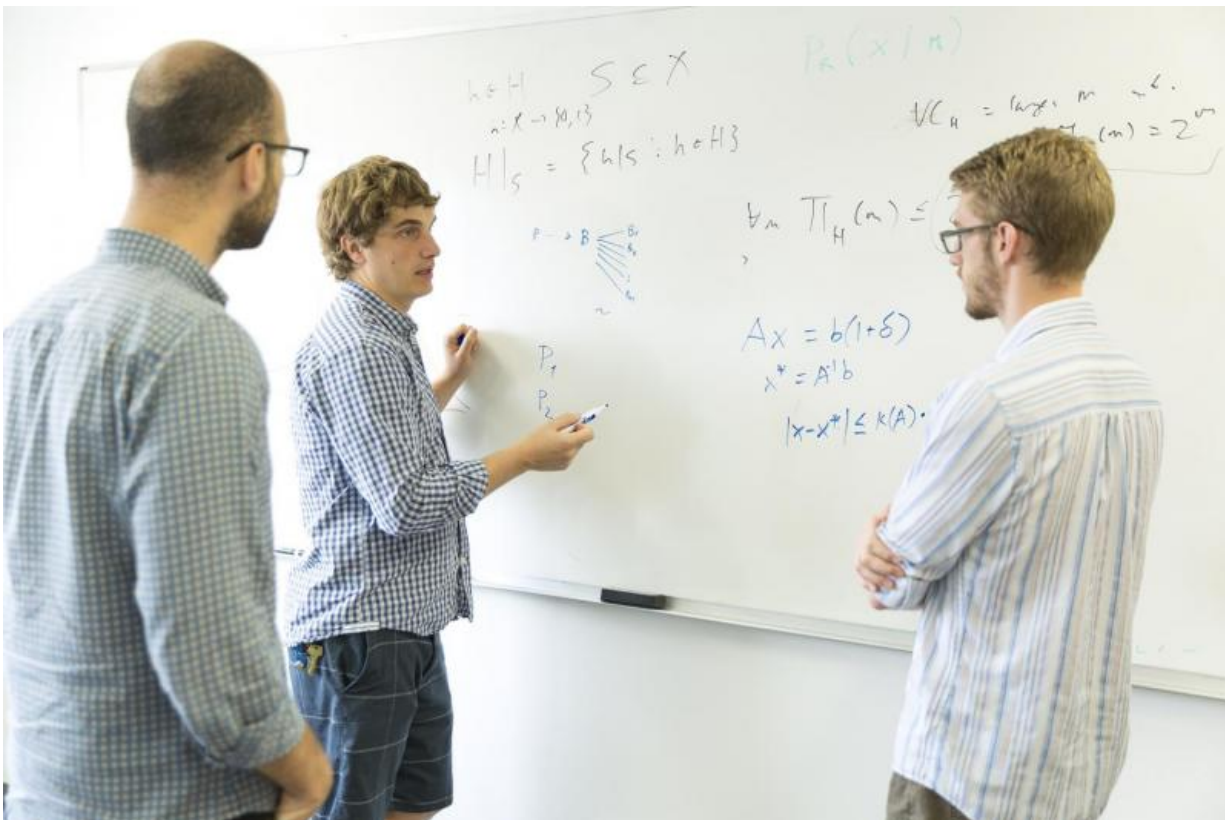# Researchers tackle bias in algorithms

July 4 2017, by Jennifer Smith

Aws Albarghouthi, Loris D'Antoni and David Merrell (left to right) are part of a team in the UW–Madison Department of Computer Sciences developing a tool to root out bias in algorithms. Credit: Sarah Morton

If you've ever applied for a loan or checked your credit score, algorithms have played a role in your life. These mathematical models allow computers to use data to predict many things—who is likely to pay back

a loan, who may be a suitable employee, or whether a person who has broken the law is likely to reoffend, to name just a few examples.

Yet while some may assume that computers remove human bias from decision-making, research has shown that is not true. Biases on the part of those designing algorithms, as well as biases in the data used by an algorithm, can introduce human prejudices into a situation. A seemingly neutral process becomes fraught with complications.

For the past year, University of Wisconsin–Madison faculty in the Department of Computer Sciences have been working on tools to address unfairness in algorithms. Now, a $1 million grant from the National Science Foundation will accelerate their efforts. Their project, "Formal Methods for Program Fairness," is funded through NSF's Software and Hardware Foundations program.

UW-Madison computer science professors Aws Albarghouthi, Shuchi Chawla, Loris D'Antoni and Jerry Zhu are leading the development of a tool called FairSquare. Computer sciences graduate students Samuel Drews and David Merrell are also involved.

What sets FairSquare apart is that it will not only detect bias, but also employ automated solutions. "Ultimately, we'd like this to be a regulatory tool when you're deploying an algorithm making sensitive decisions. You can verify it's indeed fair, and then fix it if it's not," says Albarghouthi.

Decision-making algorithms can be mysterious even to those who use them, say the researchers, making a tool like FairSquare necessary.

For example, consider a bank that uses a third-party tool to evaluate who qualifies for a mortgage or small business loan, and at what interest rate. The bank may not know how the software is classifying potential

customers, how accurate its predictions truly are, or whether results reflect racial or other types of bias.

"Many companies using these algorithms don't understand what (the algorithms) are doing," says Albarghouthi. "An algorithm seems to work for them, so they use it, but usually there is no feedback or explainability" on how exactly it is working. That makes these algorithms difficult to regulate in terms of avoiding illegal bias, he says.

Companies designing and selling these products are typically not eager to share their proprietary knowledge, making their algorithms what are known as "black box."

Says D'Antoni, "We're trying to give people the ability to ask about behaviors of an algorithm. Does it prefer a certain gender, or certain behaviors, for example?"

The stakes behind these algorithms can be high, as journalists have noted.

In a 2016 story by the investigative journalism organization ProPublica, a team of reporters examined a product used in law enforcement to predict offenders' likelihood of reoffending. The reporters uncovered troubling racial bias, though the software company in question disputes their conclusions.

According to ProPublica, "(B)lacks are almost twice as likely as whites to be labeled a higher risk but not actually reoffend." With white offenders, the opposite mistake occurred. Whites were much more likely than blacks to be pegged as low-risk yet go on to commit additional crimes.

The UW researchers are attacking the problem by isolating fairness as a

property of a software program that must be formally defined and proven.

This points to additional questions, says Drews. "Who decides what's fair? How can you be certain you're coming up with a mathematical formula that means the thing you want it to prove?"

The FairSquare team is making connections with UW–Madison scholars in other fields who can help illuminate certain aspects of this research, such as legal and ethical ramifications.

"Computing is so much more involved in people's lives these days," says Drews, making the development of FairSquare not only a significant computing challenge but also one with far-reaching social impact.

Adds Merrell, "Machine learning algorithms have become very commonplace, but they aren't always used in responsible ways. I hope our research will help engineers build safe, reliable and ethical systems."

Provided by University of Wisconsin-Madison